

DIALOG(R)File 347:JAPIO
(c) 2006 JPO & JAPIO. All rts. reserv.

07334750 **Image available**
IMAGE PROCESSING METHOD FOR DETECTING HUMAN FIGURE IN DIGITAL IMAGE

PUB. NO.: 2002-203239 [JP 2002203239 A]
PUBLISHED: July 19, 2002 (20020719)
INVENTOR(s): LUO JIEBO
APPLICANT(s): EASTMAN KODAK CO
APPL. NO.: 2001-350755 [JP 2001350755]
FILED: November 15, 2001 (20011115)
PRIORITY: 00 737026 [US 2000737026], US (United States of America),
 December 14, 2000 (20001214)
INTL CLASS: G06T-007/00; G06T-001/00

ABSTRACT

PROBLEM TO BE SOLVED: To provide a more efficient algorithm to detect comprehensive human figures in an image without assuming a pose, posture, cropping, hidden line, clothes or other atypical situations.

SOLUTION: A digital image processing method for detecting human figures in a digital color image includes the steps of providing a digital color image having pixels representing RGB values, segmenting the digital color image into nonoverlapping regions of homogeneous color or texture, detecting candidate regions of human skin color, detecting candidate regions of human faces, and for each candidate face region, constructing a human figure by grouping regions in the vicinity of the face region according to a pre-defined graphical model of the human figure, giving priority to human skin color regions.

COPYRIGHT: (C)2002,JPO

(19)日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開2002-203239

(P2002-203239A)

(43)公開日 平成14年7月19日(2002.7.19)

(51)Int.Cl. ⁷	識別記号	F I	テーマコード(参考)
G 0 6 T 7/00	2 0 0	G 0 6 T 7/00	2 0 0 Z 5 B 0 5 7
	1 0 0		1 0 0 C 5 L 0 9 6
	3 0 0		3 0 0 D
1/00	3 4 0	1/00	3 4 0 A

審査請求 未請求 請求項の数3 O L (全 10 頁)

(21)出願番号 特願2001-350755(P2001-350755)

(22)出願日 平成13年11月15日(2001.11.15)

(31)優先権主張番号 7 3 7 0 2 6

(32)優先日 平成12年12月14日(2000.12.14)

(33)優先権主張国 米国 (U S)

(71)出願人 590000846

イーストマン コダック カンパニー
アメリカ合衆国, ニューヨーク14650, ロ
チェスター, ステイト ストリート343

(72)発明者 ジェボ ルオ

アメリカ合衆国 ニューヨーク 14620
ロチェスター ユニバーシティ・パーク
805

(74)代理人 100070150

弁理士 伊東 忠彦

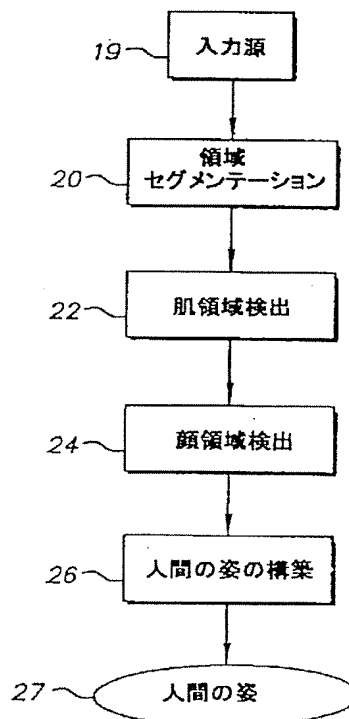
最終頁に続く

(54)【発明の名称】 デジタル画像中で人間の姿を検出する画像処理方法

(57)【要約】

【課題】 ポーズ、姿勢、クロッピング、隠線、衣装、又は他の異常な状況についての仮定をすることなく画像中の包括的な人間の姿を検出するためのより効率的なアルゴリズムを提供することを目的とする。

【解決手段】 デジタルカラー画像中の人間の姿を検出するデジタル画像処理方法は、R G B 値を表わす画素を有するデジタルカラー画像を与える段階と、デジタルカラー画像を均質な色又はテクスチャの重なり合わない領域へセグメンテーションする段階と、人間の肌の色の候補領域を検出する段階と、人間の顔の候補領域を検出する段階と、各候補顔領域について、人間の肌の色の領域に優先度を与えて、予め定義された人間の姿のグラフィックモデルに従って顔領域の近傍の領域をグループ化することにより人間の姿を構築する段階とを含む。



【特許請求の範囲】

【請求項 1】 デジタルカラー画像中の人間の姿を検出するデジタル画像処理方法であって、RGB 値を表わす画素を有するデジタルカラー画像を与える段階と、

上記デジタルカラー画像を均質な色又はテクスチャの重なり合わない領域へセグメンテーションする段階と、人間の肌の色の候補領域を検出する段階と、人間の顔の候補領域を検出する段階と、

各候補顔領域について、人間の肌の色の領域に優先度を与えて、予め定義された人間の姿のグラフィックモデルに従って上記顔領域の近傍の領域をグループ化することにより人間の姿を構築する段階とを含む方法。

【請求項 2】 上記顔領域を検出する段階は、肌色領域中の強度谷特徴の形状及び量をマッチングすることにより、人間の顔を表わす所定の強度パターンについて探索する段階を含む、請求項 1 記載の方法。

【請求項 3】 RGB 値を表わす画素を有するデジタルカラー画像中で人間の姿を検出するデジタル画像処理方法であって、

RGB 値を表わす画素を有するデジタルカラー画像を与える段階と、

上記デジタルカラー画像を均質な色又はテクスチャの重なり合わない領域へセグメンテーションする段階と、候補顔領域としての領域を選択する段階と、予め定義された人間の姿のグラフィックなモデルに従って近傍の領域をグループ化することにより人間の姿を構築しようと試みる段階と、

上記構築された姿が所定の規準を満たす場合は上記グループ化された領域を人間の姿として識別する段階とを含む方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】 本発明は概してデジタル画像処理及びコンピュータビジョンの分野に係り、更に特定のにはデジタル画像中の人間の姿を検出する方法に関する。

【0002】

【従来の技術】 デジタル画像は、画像通信において広く使用されている。デジタル画像により、画像中の関心となる対象を検出するためにコンピュータビジョン技術を使用することが可能となる。画像中に見いだされる全ての考えられる対象のうち、人間の姿は最も関心の高いものとして目立つものである。

【0003】 人間の顔検出については過去 20 年間にわたって広い研究・開発活動がなされてきた。例えば、Lobo に対して 1998 年 11 月 11 日に発行された米国特許第 5,835,616 号によって知られるものがある。Lobo の文献では、(例えばハンドヘルド式デジタルカメラ及びカムコーダといったデジタルビデオカメ

ラによって撮影され) 電子的に 2 値化された画像中の人間の顔を自動的に見つけるため、及び、顔の特徴を調べることにより顔の存在を確認するための 2 ステップ方法が開示されている。ステップ 1 は、人間の顔を検出するものであり、しわといった固有の顔の特徴や顔の画像の湾曲をよりよく示すためにぼけフィルタやエッジ強調器によってデジタル画像を強調する段階において行われる。予備フィルタリングの後、スネークレットと称されることもある予め選択された曲線が画像上に置かれ、顔の画像の自然のしわ及び曲線に合わされる。ステップ 2 は、7 つの段階で人間の顔の存在を確認するものであり、画像中の顎、顔の側面、頭部の実質的な頂部、目、口、鼻を含むデジタル画像の顔の特徴を見つけてることによって行われる。見いだされたこれらの顔特徴の間の距離の比率を、認識のために予め記憶した基準比率と比較する。

【0004】 画像の中の顔特徴を検出するためのこの方法は、エンターテインメント/アミューズメントパークの局から人口年齢統計を収集するためやテレビジョンネットワークの視聴者のレーティング調査のための人間の顔の検出といった用途に使用されうるが、これらに限られるものではない。このような収集は、客の数を数えること、ある年齢及び性別のグループを区別すること、及び/又は、特定の人々を識別することを含む。このような可能性を有するコンピュータビジョンは、更に、自動化された調査システム、人口統計調査、オートフォーカスカメラ用の自動化された写真技術、及び、人間とコンピュータの対話といった分野に適用されうる。自動化された写真技術は、合焦された被写体がないことによって低い品質を生じさせる手動調節の問題を排除する。コンピュータシステムは、ユーザの特定の要求を認識しこれに応答し、更に人間であるユーザ用に変換するためにこのシステムを使用しうる。

【0005】 種々の適用において「顔検出」が価値のあるものであることは既に知られている。しかしながら、「人物検出」又は「人間の姿の検出」は、人物検出は単に顔だけ以上のことを含むこと、及び、人物検出は顔検出がうまくいかない場合でもうまくいくことがあること、の 2 点から潜在的に更に多くの情報を与える可能性がある。

【0006】 主被写体検出、被写体についての露出補償、及び、画像合成は全て、単に顔領域だけでなく人物領域についての知識があればそれにより利益を得る。人物の写真では、主被写体は通常は、顔だけでなく、人物全体である。デジタル編集では、合成、又は、ズーム及びクロッピングを行う場合に、体の部分から切り離された顔を用いて作業するのではなく、人物全体を一体として取り扱おうとすることは理にかなっている。そして、露出の補償を行う場合は、被写体に対する正しい補償は、顔だけでなく、顔に関連する毛髪及び衣服も考慮

に入れるべきであると言える。

【0007】顔検出は、写真の中の顔が「小さすぎる」場合、例えば両目の間の距離が10画素程度であるか焦点が合っていない場合に、顔を検出するのに失敗すると予想される。このような写真の場合でも、でも幾つかの種類的人物検出器は成功しうる。この場合、人物検出は、フレームの向きの決定及び主被写体検出といった顔の検出が有効な適用において、顔検出を置き換えるものであると考えられうる。

【0008】理想的な人物検出器は、画像の各画素を、それが人物の一部であるか否かによって、また、そうであればどの人物が関連付けられるかによってラベル付けする。人物に関連付けられる画素は、体と毛髪、及び、着けている衣服、即ち、基本的に人物と共に一体として動く全てのものを含む。人物の検出は、ポーズ、姿勢、クロッピング、隠線、衣装、又は他の異常な状況にかかわらず成功せねばならない。傘、バッグ、赤ん坊といった手で保持されている物体はあいまいな領域であり、特定の用途に依存して含まれることも除かれることもある。

【0009】これは難しい問題であることが理解されるべきである。これは、1人又は2人の「肩から上の」クローズアップのビューから、座っている人物が立っている人物を部分的に隠しているような中位の距離の集合写真、殆どが隠されておりカメラに背中を向けている可能性が高い多くの人物からなる遠い群衆までを含む。

【0010】以下の文献は、従来技術で知られており同様の問題について扱う幾つかのアプローチについて記載している。

【0011】Oren外は、Proceedings of Computer Vision and Pattern Recognition, 1997においてウェーブレットに基づくテンプレートをを用いた歩行者検出の方法を開示している。この方法は、実際の対象（「歩行者」）の全ての場所及び考えられうるサイズについて画像全体に亘って所定の強度パターン（「テンプレート」）を適用するテンプレートマッチングに基づく。ウェーブレットテンプレートは、被写体の衣服及び照明条件の変化に対する感度を減少させるために使用される。この方法は、「歩行者」、即ち、歩く動きをしている詳細の少ない姿にのみ適している。この方法では、全ての場所及び寸法を網羅する探索が行われるため計算費用がかかる。

【0012】Forsyth外は、肌検出及び手足のグループ化を用いた「裸の人物」の検出の方法を開示している

(David Forsyth, Margaret Fleck, and Chris Bregler, "Finding Naked People". 1996 European Conference on Computer Vision, Volume II, pp. 592-602)。この方法では、まず、大きい面積の肌色領域を含む画像を見つけ、次に細長い領域を見つけ、これらを考えられうる人間の手足及び接続された手足のグループへとグループ化する。この方法で仮定されることは、

- ・人間は比較的簡単な形状の部分から作られること、
- ・これらの部分を組み立てるためには幾つかの方法しかないこと、
- ・組立の運動学は、多くの形態が不可能であることを補償すること、
- ・動きを測定することができるとき、これらの部分の動力学は限られていることである。

【0013】この方法が使用するモデルでは、肌領域はテクスチャを有さず、限られた範囲の色相及び飽和度を有し、

・簡単なグループ（胴体セグメント）を複雑なグループ（手足セグメント・ガードル）へ組み立てるためのグループ化規則は、3次元導体部分に対する幾何学的制約及び運動学的制約によって生ずる2次元特徴の相対的な位置についての制約を組み込み、

・エッジセグメント、即ち、略平行な局所的な対称である対のエッジ点であり間に他のエッジがないものに対してグループ化を行い、点の集合がおおよそ直線的な軸を有する領域（「リボン」）を形成し、

・手足を形成するよう、端部が近接して存在し、断面の長さが略同じであるリボンの対をグループ化し、

・手足を推定される帯へグループ化し、

・脊髄・腿のグループを形成するようセグメントをグループ化する。

【0014】かかる方法では、

- ・提案されるグループ化規則が完全でないこと、及び、
- ・衣服には複雑な模様が施されている場合が多く、表面の向きの変化によって生ずる歪みを受けるため、着衣した人物をセグメンテーションすることが困難であること、という問題点がある。

【0015】Felzenszwalb及びHuttenlocherは、Proceedings of Computer Vision and Pattern Recognition, 2000の中で、変形可能なモデルを用いた人間の姿のマッチングのための方法を開示している。この方法は、部分の対の間のバネ状の接続によって表わされる変形可能なモデルのマッチングに基づく。人間の姿のモデルは以下の通りであり、各部分は同じ強度の矩形のボックスに当てはまり、各接続は相対的な向き、寸法、及び節の整列の夫々の偏りに関連付けられる費用を特定する。

【0016】この方法の問題点は、

- ・検出は既知の人間の姿を既知の部分にマッチングする場合に限られていること（画像中の予想される人物に対してモデルが構築されねばならないこと）、
- ・部分が頑強な方法で得られないこと（考えられうる部分の位置は、x位置及びy位置のそれぞれについて50バケットへ、寸法について10バケットへ、向きについて20バケットへ量子化される）、
- ・マッチングの結果は、人物のおおよその位置であり、多くの適用には適していないことである。

【0017】Oren外による方法は、人間の姿が、立った

状態（歩行者）であり、隠線がないか又はわずかである全身の姿であると想定する。

【0018】Forsyth外によって教示される方法は、以下のような多数の仮定に基づくものであり、即ち、

- ・人間の姿が裸であること、
- ・全ての人間の体の部分は肌領域として検出されうること、
- ・全ての背景領域は肌領域としては検出されないこと、を仮定する。

【0019】Felzenszwalb外によって教示される方法は、未知の人間の姿の検出というよりもむしろ、主に既知の人間の姿のマッチングのために設計されたものである。この方法は、例えば以下の仮定、即ち、

- ・人間の姿のモデルは予め特定され、変化しないこと、
- ・人間の体の全ての露出した部分が均一な肌色の領域として検出されうること、及び、
- ・全ての衣服の部分が均一に色づけされた領域として検出されること、という仮定に基づく。

【0020】しかしながら、上述の仮定は多くの画像適用に対しては成り立たない。例えば、多くの適用において、探索の前に人間の姿のモデルを構築すること、又は姿勢を立った位置に制限することは可能でない。また、殆どの適用では、人間はなんらかの衣服を着用している。

【0021】

【発明が解決しようとする課題】従って、ポーズ、姿勢、クロッピング、隠線、衣装、又は他の異常な状況についての仮定をすることなく画像中の包括的な人間の姿を検出するためのより効率的なアルゴリズムについての要請がある。画像がある程度の品質であり異なる領域を識別することができること、及び、人間の姿が適度な寸法であり胴体部がセグメンテーションされうるといふことのみが仮定される。

【0022】

【課題を解決するための手段】本発明によれば、従来技術の問題に対する解決策が提供される。これは、本発明によれば、RGB値を表わす画素を有するデジタルカラー画像中で人間の姿を検出するデジタル画像処理方法であって、画像を均質な色又はテクスチャの重なり合わない領域へセグメンテーションする段階と、人間の肌色の候補領域を検出する段階と、人間の顔の候補領域を検出する段階と、各候補顔領域について、人間の肌の色の領域に優先度を与えて、予め定義された人間の姿のグラフィックモデルに従って顔領域の近傍の領域をグループ化することにより人間の姿を構築する段階とを含む方法によって達成される。

【0023】本発明の特徴によれば、RGB値を表わす画素を有するデジタルカラー画像中の人間の姿を検出するデジタル画像処理方法であって、RGB値を表わす画素を有するデジタルカラー画像を与える段階と、

デジタルカラー画像を均質な色又はテクスチャの重なり合わない領域へセグメンテーションする段階と、人間の肌色の候補領域を検出する段階と、人間の顔の候補領域を検出する段階と、各候補顔領域について、人間の肌の色の領域に優先度を与えて、予め定義された人間の姿のグラフィックモデルに従って顔領域の近傍の領域をグループ化することにより人間の姿を構築する段階とを含む方法が提供される。

【0024】

【発明の実施の形態】以下の説明では、本発明の望ましい実施例について通常はソフトウェアプログラムとして具現化される画像処理方法として説明する。当業者はかかるソフトウェアと同等のものがハードウェア中に構築されうること容易に認識するであろう。画像操作アルゴリズム及びかかるアルゴリズムを実行するシステムは周知であるため、本願では特に本発明の方法の部分を形成する、又はより直接的に協働する、アルゴリズム及びシステムについて説明するものとする。本願で特に図示、説明、又は提案されていないかかるアルゴリズム及びシステム、並びに、画像信号を生成及び他の方法で処理するハードウェア及び／又はソフトウェアの他の面は、従来技術で知られているかかるシステム、アルゴリズム、構成要素、及び要素から選択されうる。以下の記載にあるように本発明によって説明される方法が与えられていれば、本発明の実施に有用な本願に特に図示又は記載されていないソフトウェアは従来通りであり当業者の知識の範囲に含まれる。

【0025】更に、本願では、コンピュータプログラム製品は、例えば、磁気ディスク（例えばフロッピー（登録商標）ディスク）又は磁気ディスクといった磁気記憶媒体、光ディスク、光テープ、又は機械読み取り可能なバーコードといった光学記憶媒体、ランダムアクセスメモリ（RAM）又は読み出し専用メモリ（ROM）といった固体電子記憶装置又は本発明による方法を実施するための命令を有するコンピュータプログラムを記憶するために使用される任意の他の物理的な装置又は媒体といったコンピュータ読み取り可能な記憶媒体に記憶されうる。

【0026】本発明は、デジタル画像中で人間の姿の位置を探す方法に関する。図1を参照するに、本発明を実施するのに適した典型的な画像処理システムが示される。入力デジタル画像は、例えば、デジタルカメラ又はスキャナ、インターネット接続、別の記憶装置等といった入力源10から入来する。次に、コンピュータ12は、入力デジタル画像を本発明に従って処理し、例えばデジタルプリンタ、表示装置、インターネット接続、別の記憶装置等といった出力装置14へ送信される出力画像を生成する。

【0027】図2は、以下詳述する本発明による画像処理方法の現在の実施例を示すフローチャートである。

【0028】図2に概略的に示されるように、本発明の画像処理方法は、人間の姿の出力27を発生するために入力画像19を処理するための4つのモジュール、即ち、画像セグメンテーションモジュール(20)、肌検出モジュール(22)、顔検出モジュール(24)、及び、人間の姿構築モジュール(26)である。以下、各モジュールについて個別に詳述する。

【0029】・画像セグメンテーションモジュール(20)

任意の形状の物理的に整合した追跡可能な数を発生させるために、適応ベイズ色セグメンテーションアルゴリズム(Luo et al., "Towards physics-based segmentation of photographic color images," Proceedings of the IEEE International Conference on Image Processing, 1997)が使用される。このセグメンテーション方法を用いることが望ましいが、当業者は本発明の範囲を逸脱することなく任意の形状の対象領域を得るために異なるセグメンテーション方法を使用しうることを認識するであろう。任意の形状の領域のセグメンテーションは、

- (1) 対象の寸法、形状、位置と対象間の空間的な関係の正確な尺度、
- (2) 対象の色及びテクスチャの正確な尺度、並びに、
- (3) 主要な被写体の正確な分類という利点を提供する。

【0030】図4を参照するに、望ましいセグメンテーションアルゴリズムのブロック図が示される。まず、画像から領域への最初のセグメンテーションが行われ(50)。セグメンテーションは、画像の色ヒストグラムを集め、ヒストグラムを画像中の別個の目立った色に対応する複数のクラスタへ分割することによって行われる。

(Luo et al., "Towards physics-based segmentation of photographic color images," Proceedings of the IEEE International Conference on Image Processing, 1997)に記載されるように、色クラスタの平均値に対する望ましい物理量に基づく色距離メトリックに従って、各画素は色空間中の最も近いクラスタへ分類される。この分類処理により画像の初期セグメンテーションが行われる。各画素には近傍ウィンドウが配置され、これはこの画素のローカルな色ヒストグラムを計算するためにどの近傍画素が使用するかを決定するために用いられる。ウィンドウの大きさは、ローカルな色ヒストグラムが全体画像の色ヒストグラムと同じであり再計算される必要がないよう、最初は画像全体の大きさに設定される。

【0031】次に、2つの交互の処理、即ち現在のセグメンテーションに基づいて各色クラスのローカルな平均値を再計算する処理(54)と、色のクラスの更新されたローカルな平均値に従って画素を再び分類する処理

(56)との反復的な手順が実行される。この反復的な手順は収束に達するまで実行される(60)。この反復的な手順中、空間的な制約の強さは徐々に調整されうる

(58)(例えば、空間的な制約の強さを示す β の値は各繰り返しの都度、線形に増加する)。特定のウィンドウの大きさについて収束に達した後、色クラスのローカルな平均値を推定するために使用されたウィンドウは大きさが半分に減少される(62)。この反復的な手順は、色クラスについてのローカルな平均値がより正確に推定されることを可能とすべく、減少されたウィンドウの大きさについて繰り返される。この機構は、セグメンテーション処理に空間的な適応性を導入する。最終的に、反復的な手順が最小のウィンドウのサイズについての収束に達したときに画像のセグメンテーションが行われる(64)。

【0032】上述のセグメンテーションアルゴリズムは、テクスチャのセグメンテーションを行うために拡張されうる。セグメンテーションへの入力として色値を用いる代わりに、同じ枠組みを用いてテクスチャのセグメンテーションを行う場合にはテクスチャ特徴が用いられる。テクスチャ特徴の種類の例としては、ウェーブレット特徴がある(R. Porter and N. Canagarajah, "A robust automatic clustering scheme for image segmentation using wavelets," IEEE Transaction on Image Processing, vol. 5-, pp. 662-665, April 1996)。

【0033】更に、色とテクスチャの特徴を組み合わせたものに基づいて画像のセグメンテーションを行うために、図4の方法への入力として色値及びウェーブレット特徴から構成される組み合わせられた入力を使用される。組み合わせられた色及びテクスチャのセグメンテーションにより、均質な色又はテクスチャのセグメンテーションされた領域が得られる。

【0034】肌検出の主な目的は、腕、手、脚、足といった潜在的に露出した人体の部分の識別することである。顔検出の主な目的は、(顔を識別する)という明らかなものに加え、人間の姿の構築における人類学的な部分の制約及び運動学的な動きが強められるよう、尺度のおおよその推定を与えることである。

【0035】肌検出の他の目的は、グラフィックモデルから顔を含まないであろう領域に一致するものを排除することによって計算上の費用を減少させることである。このような排除された領域は更には分析されないため、この段階は、擬陽性の確率が高くなるという犠牲を払ってでも、擬陰性の確率が非常に低くなるよう調整されねばならない。

【0036】同様に、顔検出の他の目的は、グラフィックモデルから顔でない確率の高い領域に一致するものを排除することによって計算上の費用を減少させることである。このような排除された領域は更には分析されないため、擬陽性の確率が高くなるという犠牲を払ってでも、擬陰性の確率が非常に低くなるよう調整されねばならない。

【0037】・肌領域検出(22)

本発明で用いられる肌検出アルゴリズムは、カラー画像セグメンテーションと、望ましい色メトリック Lst における所定の肌分布を用いる (Lee, "Color image quantization based on physics and psychophysics," Journal of Society of Photographic Science and Technology of Japan, Vol. 59, No.1, pp. 212-225, 1996)。肌領域は、セグメンテーションされた領域の平均色の分類を行うことによって獲得されうる。確率値は、人間の姿を構築する段階が二値決定の代わりに確率を必要とする場合にも保持されうる。

【0038】肌検出方法は、輝度成分及びクロミナンス成分における人間の肌の色の分布に基づく。概して、RGB画素値のカラー画像は望ましい Lst メトリックへ変換される。すると、3次元ヒストグラムが形成され平滑化される。次に、3次元ヒストグラムのピークの位置が探され、ヒストグラムの各欄にピークを割り当てることにより欄のクラスタリングが行われる。各画素は、画素の色に対応する欄に基づいて分類される。人間の肌の平均的な色 (Lst) 値と接続された領域の平均の色に基づいて、肌確率が計算され、確率が所定の閾値よりも

【0039】・顔領域検出 (24)

顔アルゴリズムは、候補肌領域中の主な顔特徴 (目、眉毛、鼻、及び、口) の検出に基づいて潜在的な顔を識別する。

【0040】本発明においては原始的な顔検出アルゴリズム (図5) が使用される。これは、肌検出段階によって出力された肌マップを他の顔に関連するヒューリスティックと組み合わせることにより、画像中の顔の位置に関する信念を出力する。肌領域 (72) として識別される画像中の各領域 (70) は、楕円に当てはめられる

(42)。楕円の長軸及び短軸は、やはり楕円の外側領域の画素の数及び楕円の中の領域の一部でない部分の画素の数として計算される。縦横比は、長軸対短軸の比率として計算される。顔の確率は、当てはめられた楕円の縦横比と、楕円の外側の領域の面積と、楕円の中の領域の一部でない面積との関数である。再び、特定の領域が顔であるか否かについての2値決定を生じさせるため確率値は保持されるか単に所定の閾値と比較されうる。

【0041】更に、候補顔領域中のテクスチャは、顔である尤度を更に特徴付けるために使用されうる。谷検出 (76) は、顔特徴 (目、鼻孔、眉毛、及び、口) が存在する可能性の高い谷部を識別するために使用される。この処理は、顔領域 (78) から顔でない肌領域を分けるために必要である。

【0042】谷検出は、画像の輝度信号中で行われる。これには一連のグレースケール形態演算が含まれる。形態カーネルは、最善の結果のためには半球状であることが望ましいが、より高速の分離可能な1次元フィルタリング演算を容易とするために矩形のカーネルであっても

よい。谷検出は以下のものを含む。

【0043】・形態学的な谷検出

次に、画像中の均質な領域及び明るい領域を除去するため、グレースケール画像に対して形態的なTOP-HAT (一) 操作 (典型的には要素サイズ=5) が適用される (Gonzalez and Woods, Digital Image Processing, Addison-Wesley, 1992, Chapter 8.4参照)。

【0044】・雑音抑制

孤立した衝撃雑音画素を除去するため、先行する段階から結果として得られる画像に対してメディアンフィルタリング (典型的には要素サイズ=3) が適用される。

【0045】この場合、顔の確率は、当てはめられた楕円の縦横比と、楕円の外側の領域の面積と、楕円の中の領域の一部でない面積と、領域中の谷画素の割合の関数である。

【0046】当業者は、本発明を逸脱することなく異なる顔検出方法を使用することが理解されるであろう。実際、図3(a)を参照するに、顔は、明らかに人間の姿のモデルの「ネットワーク」の「ハブ」である。信頼性の高い顔検出は、(1) 実行されねばならない探索の範囲を狭め、(2) 人間の姿のモデルにやはり当てはまる領域のグループの偶然一致する形態による擬陽性の数を減少させる。このような偶然一致する形態は稀であると予想すると考えるのは妥当であり、従って、人間の姿の構築の次の段階はロバストな検出段階よりも低い検出段階又は顔検出段階なしで初期化されうる。そのような場合、より多くの探索を実行する必要がある、擬陽性の確率が高くなることを予想すべきである。例えば、顔検出を行わない場合、各領域が潜在的な顔領域であり、非常に多数の探索がされると仮定せねばならない。

【0047】・人間の姿の構築 (26)

ここで、全ての部分が抽出されているため、異なるポーズ及びピクチャフレーミングに対応する可能な人間の姿の形態に一致するように正しい部分が組み合わせられれば人間の姿が検出されうる。ピクチャフレーミングは、ピクチャが実際のシーンをどのように捕捉するかについて参照される。各ピクチャは境界を有し、一方、それに対応する実際のシーンは境界を有さない。従って、例えば人間の姿といった対象のある部分がクロッピングされデジタル画像中に現れない可能性もある。一般的には、人間の姿はピクチャフレーミング中に上側ではなく下側でクロッピングされる可能性が高い。

【0048】図3(a)を参照するに、意味ネットワークを用いて表わされた人物の全身のグラフィックモデルが示される。ノードは体の部分であり、有向リンクは人類学的な比率及び運動学的な動きから制約を符号化する。リンクは、ノードから他のノードへ、又はそのノード自身を指しうる。前者は関係を符号化するものであり、後者は性質を符号化するものである。

【0049】顔ノード101は、ネットワークの「ハ

ブ」又は「焦点」であり、なぜならば本願発明では残る体の部分のスケールの推定値を導出するためにこれを使用するからである。

【0050】髪ノード102は、顔ノード101に直接的にリンクされているだけであり、顔に対して妥当な大きさでなくてはならない（例えば顔の大きさに対して10%乃至100%）。実際、全ての体の部分は顔に対して妥当な大きさでなくてはならず、これらの制約は自分を指すリンクによって符号化される（図4には明らかに示されていない）。更に、髪の色及びテクスチャは大幅に変化しうるが、一般的な髪の色は限られている（黒髪、金髪、ブロンド、赤毛、白髪である可能性は高いが、紫、青、緑である可能性は低い）。

【0051】シャツノード103は顔ノード101にリンクされる。シャツノード103は、首と首回りの服飾品（ネックレス、スカーフ等）により必ずしも顔に接続されているわけではない。シャツノード103は、顔の下にあり、妥当的な範囲内でなくてはならず（例えば、最も遠い点が顔の高さの4倍以内であり）、顔と比較して相対的に大きい寸法（例えば顔の寸法の400%乃至800%）でなくてはならない。シャツの色及びテクスチャには制約はない。更に、シャツノードは、左及び右の腕／手ノード104及び105と、ズボンノード106とにリンクされる。

【0052】腕／手ノード104及び105は、領域が肌であるという要件を強めるために自己を指すリンクを有する。腕／手は、顔の妥当的な範囲内でなくてはならず（例えば、最も遠い点が顔の高さの3倍以内であり）、顔と比較して妥当な寸法（例えば顔の寸法の25%乃至100%）でなくてはならず、シャツに直接接続されている。腕時計及びブレスレットといった小さい腕装着物の装着物が手から腕への肌領域を中断しうるが、実際上はこれらは無視される。

【0053】ズボンノード106はシャツノード103と、2つの脚ノード107及び108にリンクされる。ズボンノード106は、顔と比較して妥当な寸法（例えば顔の寸法の200%乃至600%）でなくてはならず、顔の妥当的な範囲内（例えば、最も遠い点が顔の高さの8倍以内）でなくてはならない。

【0054】2つの脚ノード107及び108はズボンノード106と2つの靴ノード109及び110に夫々リンクされる。脚ノードは、肌の要件を強めるために自己を指すリンクを有する。脚は、顔の妥当的な範囲内でなくてはならず（例えば、最も遠い点が顔の高さの8倍以内であり）、顔と比較して妥当な寸法（例えば顔の寸法の400%乃至800%）でなくてはならず、ズボンに直接接続される。靴下が脚と靴を分離する場合があるが、これらは実際上は靴の一部であると考えられる。

【0055】靴ノード109及び110は、脚ノード107、108に夫々リンクされる。靴領域は顔と比較し

て妥当な寸法（例えば顔の寸法の50%）でなくてはならない。

【0056】図3の（b）には、図3の（a）のうちノード101、102、103のみを示す簡単化されたグラフィックモデルが示されている。

【0057】考えられるポーズ及びピクチャフレーミングの他の人間の姿の形態は図6の（a）乃至（d）、並びに、図7の（a）及び（b）に示される。図6の

（a）は、単一の完全な長さの人間の画像を有する原画像である。図6の（b）はセグメンテーションの後の結果を示す図である。図6（c）は肌検出の後の結果を示す図であり、図6の（d）は個々の人物を示す図である。図7（a）は2人の人間の上半身の原画像を示す図である。図7の（b）はセグメンテーションの後の結果を示す図であり、図7の（c）は肌検出段階の後の結果を示す図であり、図7の（d）及び（e）は個人の検出を示す図であり、図7の（f）は図7の（d）と（e）を合わせたものである。

【0058】体の部分の完全性は多様である。最も完全性の低い形態（従って人間の姿の構築）から開始して、全ての生じうる形態について調べられる。最も完全な形態（例えば図3の（a））には、検出された場合は、最も高い信頼性のスコアが割り当てられる。

【0059】グラフィックなモデルとそれらのマッチングは、意味ネット、ベイズネット、及び因子グラフといった多くの方法を用いて行われうる。ここでは、意味ネットに基づく実施について説明する。意味ネットは、属性の値を対応する所定の範囲の値と比較することによって夫々が上述の制約を特定する一組の規則を用いてリンクを符号化する。各リンクは、「オン」（有効）又は「オフ」（無効）である。これに対して、ベイズネット及び因子グラフは所定の確率関数を用いてリンクを符号化する。従って、このような確率ネットワークが用いられる場合は肌及び顔の確率を保持することが望ましい。人間の姿を検出する全体的な確率は、確率ネットワークが落ち着いた後に得られる。

【0060】ベイズネット（J. Pearl, Probabilistic Reasoning in Intelligent Systems, San Francisco, CA: Morgan Kaufman, 1988）は、グラフ中の種々の実体間の因果関係を表わす有向非循環グラフである。リンクの方向は因果性を表わす。これは、種々の実体の間の評価手段を知る同時確率密度関数（PDF）である。利点としては、明示的な不確実性特徴付け、高速且つ効率的な計算、迅速な学習、高い適応性及び構築の容易さ、及び、人間の理由付けの枠組みにおいて前後関係的な知識を表わすことがある。ベイズネットは、以下の4つの構成要素からなる。

【0061】・事前確率：ベイズネット中の種々のノードに関する初期の信念

・条件付き確率行列（CPM）：ベイズネット中の2つ

の接続されたノード間の統計的な関係

・証拠：ベイズネットに入力された特徴検出器からの観察値

・事後確率：ベイズネットを通じて証拠が伝搬された後の最終的な計算された信念。

【0062】因子グラフ (F. R. Kschischang, B. J. Frey and H. -A. Loeliger, "Factor graphs and the sum-product algorithm", 1998年7月にIEEEの情報理論についての紀要に提出) は、種々のノード及び関数を有する二分グラフである。論理関数は各関数ノードに關連付けられ、隣接する変数にのみ依存する。グローバル関数はローカル関数の積として与えられる。ベイズネットは、より一般的な因子グラフへ変換されうる。

【0063】

【発明の効果】本発明によれば、デジタル画像を認識しそれにより人間に理解可能な対象、属性又は条件に対して有用な意味を割り当てるためにデジタル式に処理し、得られた結果をデジタル画像の更なる処理に用いる技術であると理解されるデジタル画像理解技術が提供される。

【0064】本発明によれば、着衣した未知の人間の姿をデジタル画像中でより高い信頼性で検出することが可能となるという利点がある。

*

* 【図面の簡単な説明】

【図1】本発明による画像処理方法を実施するのに適した画像処理システムを示すブロック図である。

【図2】本発明による画像処理方法の1つの実施例を示すフローチャートである。

【図3】人間の全身モデルと本発明による対応するグラフィックモデルとを示す図である。

【図4】人間の上半身領域と本発明による対応するグラフィックモデルとを示す図である。

【図5】本発明で使用される顔検出方法を示すブロック図である。

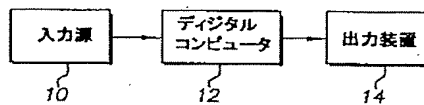
【図6】(a)乃至(d)は本発明による人間の体の検出の例を示す図である。

【図7】(a)乃至(f)は本発明による人間の体の検出の例を示す図である。

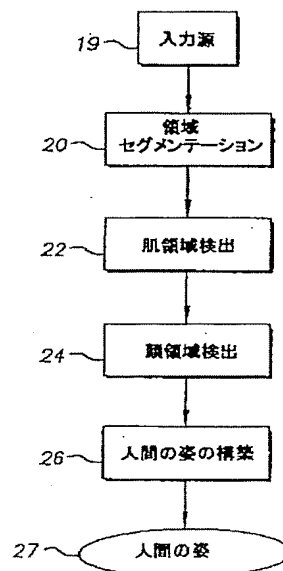
【符号の説明】

- | | |
|----|-------------|
| 19 | 入力源 |
| 20 | 領域セグメンテーション |
| 22 | 肌領域検出 |
| 24 | 顔領域検出 |
| 26 | 人間の姿の構築 |
| 27 | 人間の姿 |

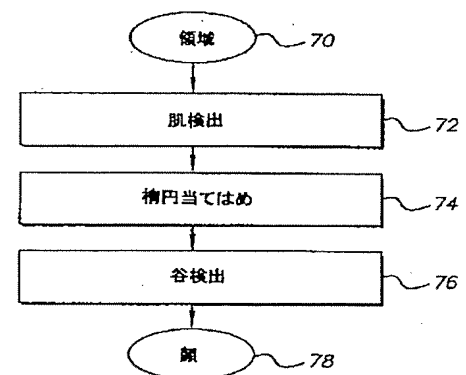
【図1】



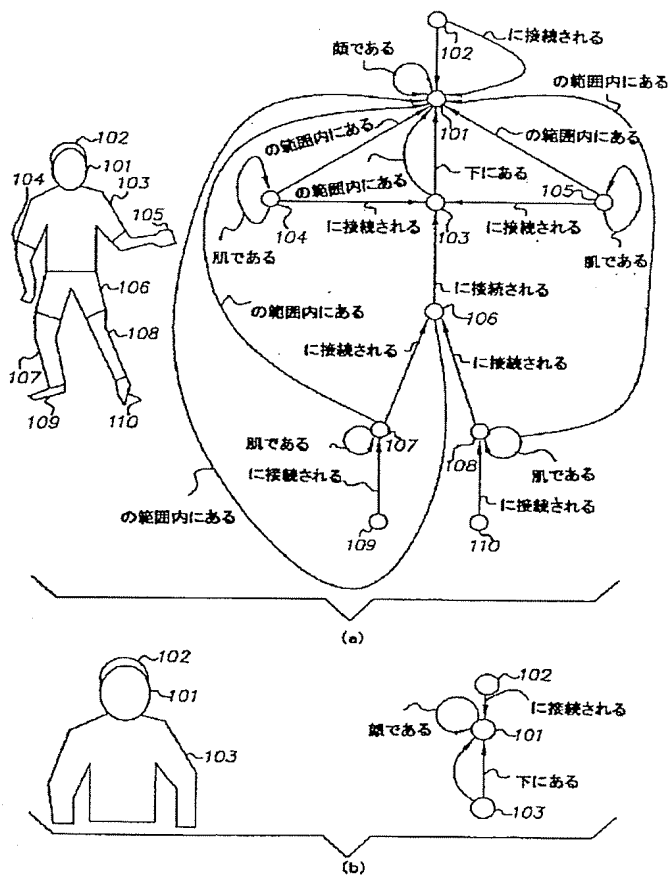
【図2】



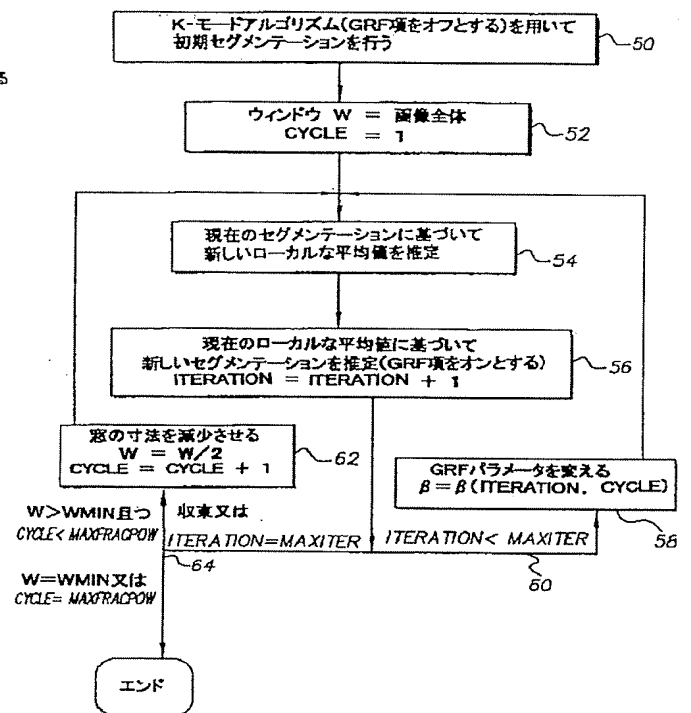
【図5】



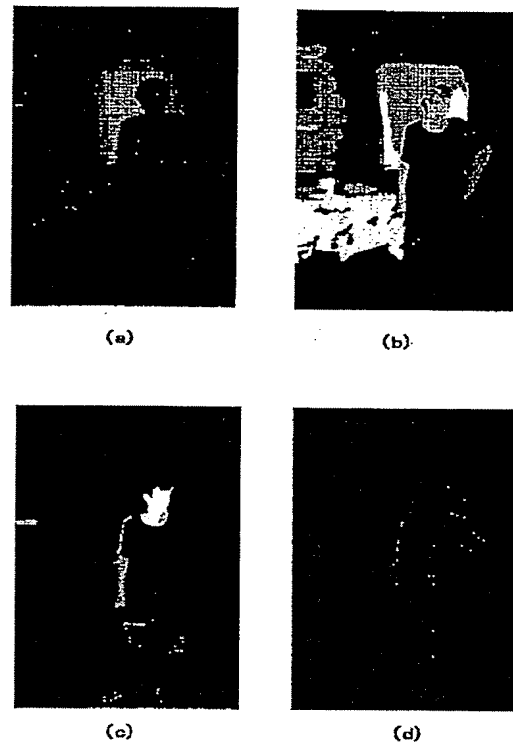
【図 3】



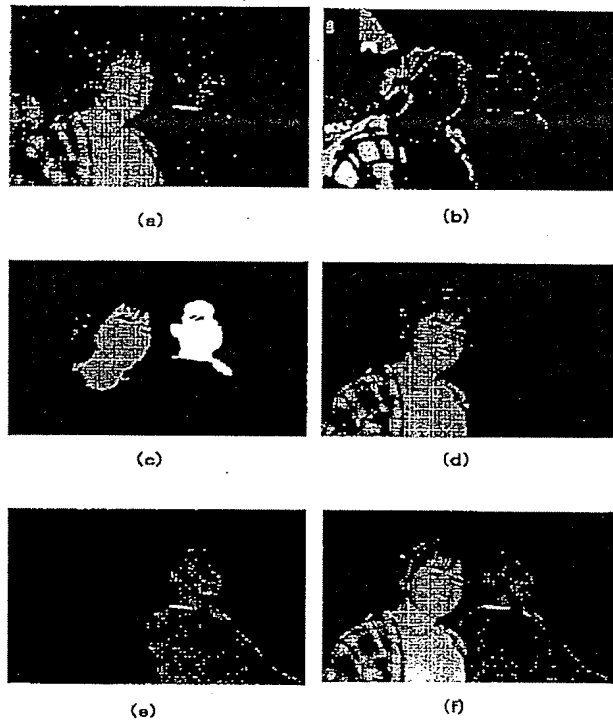
【図 4】



【図 6】



【図7】



フロントページの続き

Fターム(参考) 5B057 CA01 CA08 CA16 CB01 CB08
CB12 CB16 CE09 DA08 DB02
DB06 DC23 DC32
5L096 AA02 AA06 EA35 FA15 FA37
FA46 GA38 GA40 HA07 JA09
JA11 JA18